

# LC-MS Peak Annotation and Identification with *CAMERA*

Carsten Kuhl, Ralf Tautenhahn and Steffen Neumann

May 3, 2010

## Introduction

The R-package *CAMERA* is a (Collection of Algorithms for MEtabolite pRofile Annotation). It's primarily used for annotation of LC-MS data. Therefore it interacts directly with processed data from *xcms* and *Rdisop* for additional analyses.

It includes the annotation of isotope peaks, adducts and fragments in peak lists generated by *xcms*. Additional methods group together mass signals measured from a single metabolite, based on rules for mass differences and peak shape comparison [1].

Based on this annotation, the molecular composition can be calculated if the mass spectrometer has a high-enough accuracy for both the mass and the isotope pattern intensities.

## 1 Peak Annotation

### 1.1 Adduct list and molecular mass estimation

For soft ionisation methods such as LC/ESI-MS, different adducts (e.g.  $[M + K]^+$ ,  $[M + Na]^+$ ) and fragments (e.g.  $[M - C_3H_9N]^+$ ,  $[M + H - H_2O]^+$ ) occur. Depending on the molecule having an intrinsic charge,  $[M]^+$  may be observed as well. In most cases, substances generates a bulk of different ions, so deconvolution is necessary.

Therefore an estimation of the molecular mass of  $[M]$  can be calculated from at least two annotated adduct ions. To scan for adducts every theoretical possible combination of adducts from a given list of ions are calculated. For a small example of common ions see Tab. 1. Every group of peaks is scanned,

Formula	Mass difference in amu
$[M + H]^+$	1.007276
$[M + Na]^+$	22.98977
$[M + K]^+$	38.963708
$[2M + Na]^+$	22.98977
$[M + H + Na]^{2+}$	23.9976
...	...

Table 1: Examples of calculated adducts for the Kations (K,H,Na) with their mass differences occurring in positive ion mode. The actual difference is calculated considering the charge and the number of molecules  $M$  in the observed ion.

if these combinations fit with the mass differences and then molecular masses are computed.

## 2 Processing with *CAMERA*

### 2.1 Preprocessing with *xcms*

At first, create an *xcmsSet* with your favourite parameters, e.g.:

```
library(CAMERA)
#Single sample example
file <- system.file('mzdata/MM14.mzdata', package = "CAMERA")
xs <- xcmsSet(file,method="centWave",ppm=30,peakwidth=c(5,10))

#Multiple sample
library(faahKO)
filepath <- system.file("cdf", package = "faahKO")
xsg <- group(faahko)
```

### 2.2 Annotation Workflow

The annotation procedure can be split into two parts: first, we want to answer the questions which peaks/ions from one substance belongs together and second, compute the exact mass of the molecule. The principal use of CAMERA is demonstrated in the next sections. See the manpages for further information about the functions and their parameters .

The annotation workflow in CAMERA contains four different functions: For dealing with the first question:

1. peak grouping after retention time (`groupFWHM`)
2. peak group verification with EICs correlation (`groupCorr`)

Both methods separate the peaks into groups, which we define as "pseudospectra". These pseudospectra can consists from one up to 100 ions, depending on your molecule and solves the first task. And for the second:

1. annotation of possible isotopes (`findIsotopes`)
2. annotation of adducts and calculating hypothetical masses for the group (`findAdducts`)

This results in a data-frame similiar to a *xcms* peak table, that can be easily stored in a comma separated table (Excel-readable). The use of these methods can be in different order, which is shown in the next section.

### 2.2.1 Working with single sample

Let's come to the practical work. Here we have a single sample file either in positive or negative ionisation mode. The *xcmsSet* was created as shown in section 2.1.

```
# Create an xsAnnotate object
an  <- xsAnnotate(xs)
# Group after RT
anF <- groupFWHM(an, perfwHM = 0.6)
# Annotate Isotopes
anI <- findIsotopes(anF, mzabs = 0.01)
# Verify grouping
anIC <- groupCorr(anI, cor_eic_th = 0.75)
#Annotate adducts
anFA <- findAdducts(anIC, polarity="positive")
```

In the above example, we first create the pseudospectra with the retention-time information. The *perfwHM* parameter defines the window width, which is used for matching. Lower it for a smaller windows or set it to a higher value, if the retention time varies. This step generate 14 pseudospectra. Afterwards we annotate isotopic peaks, with *mzabs* as allowed m/z error. In this example we find 36 isotope peaks, with means the number of  $[M + 1]$ ,  $[M + 2]$ , *ldots* ions. These isotope information are useful in the next step, where for every peak in one pseudospectra a pairwise EIC correlation is done. If the correlation value between two peaks is higher than the threshold *cor\_eic\_th* it will stay in the group, otherwise both are separated. If the peaks are annotated isotope ions, they will not be divided. That separate

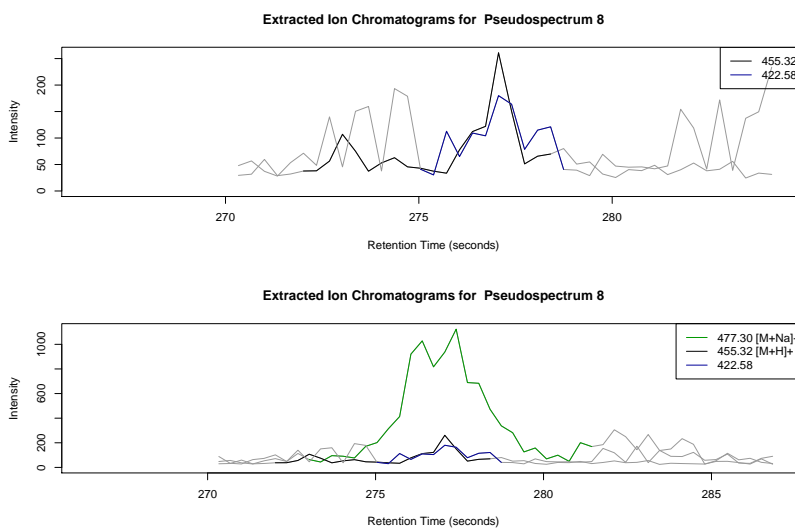


Figure 1: Example of using the polarity parameter

our 14 pseudospectra into 35. Additionally you can set an constraint, that the primary adducts e.g.  $[M + H]$  and  $[M + Na]$  stay together, even if the correlation value is small. To do so call the function with an polarity parameter like in `findAdducts` (`groupCorr (... , polarity="positive")`), which only separate the 14 into 32 groups. The difference is only small, three single ions were not sorted out, but as you can see the importance in figure ?? . In the upper example a annotation was possible in contrast to the lower one, because the  $[M + Na]$  has put into his own group. So we strongly encourage the use of the polarity mode.

After the second pseudogroup creating step we now finally do a complete annotation of adducts. Therefore the *polarity* parameter must be set. For exporting the result simply do:

```
peaklist <- getPeaklist(xsa)
write.csv(peaklist, file='xsannotated.csv')
```

where file is the ouput filename. That's so far as for the simple sample approach. Please note that every method has additional parameters, that are not explicitly mentioned here. Also if your analysis doesn't need annotations, only a separation into groups, then simply stop after `groupCorr`. The grouping results are stored in `object@pspectra`, which is a list, that saves as element the peakindexes for every group.

```
#xsa is here the result from findAdducts
```

```

peak.idx <- xsa@spectra[[1]]
#print the indexes of all peaks from pseudospectrum 1
cat(peak.idx)

```

## 2.2.2 Working with multiple sample

In this case we have multiple samples like replicates of one probe or e.g. a wildtype vs. mutant experiment. As in the example before, we start with the already processed `xcmsSet`-object. Note: If you want to do an `diffreport` later on, make sure that you run `fillPeaks` on your `xcmsSet` before.

As test dataset we use here the `faahKO`. For more information about that dataset see <http://dx.doi.org/10.1021/bi0480335>. CAMERA contains different approaches, how it deals with multiple sample datasets. Here we only show the most common way, for the other strategies see the manpage of `xsAnnotate`, especially the parameter `sample`.

```

#Create an xsAnnotate object
xsa <- xsAnnotate(xsg)
#Group after RT value of the xcms grouped peak
xsaF <- groupFWHM(xsa, perfwhm=0.6)
#Verify grouping
xsaC <- groupCorr(xsaF)
#Annotate isotopes, could be done before groupCorr
xsaFI <- findIsotopes(xsaC)
#Annotate adducts
xsaFA <- findAdducts(xsaFI, polarity="positive")

#Get final peaktable and store on harddrive
write.csv(getPeaklist(xsaFA),file="result_CAMERA.csv")

```

Similar to the previous example, the grouping is followed by an annotation. But in contrast to that, we now have more additional summaries respectively analysis functions. For a comparison and statistical analysis between different sample classes, `xcms` contains the `diffreport` function. CAMERA can use this method for better representation.

```

#Run fillPeaks on xcmsSet
xsa.fill <- fillPeaks(xsg)
#Make a diffreport with CAMERA result
diffreport <- annotateDiffreport(xsa.fill)
#Save on harddrive
write.csv(diffreport, file="diffreport.csv")

```

The `annotateDiffreport` is a wrapper for the `xcms diffreport` function and combines the results from CAMERA. The resulting table has three different columns, see section 2.3. For a speed up it's possible to preselect pseudospectra or make an automatic selection based on the `diffreport` result. For example select only groups with a fold change higher than 4.

```
#Example 1 with creating list of interest from grouped xcmsSet
diffreport <- annotateDiffreport(xsg.fill, quick=TRUE)
#save Results
write.csv(diffreport, file="diffreport.csv")
#Look into the table and select interesting pseudospectra
#e.g. pseudospectra 10,11 and 30
psg_list <- c(10,11,30)
diffreport.annotated <- annotateDiffreport(xsg.fill, psg_list=psg_list,
polarity="positive")

#Example 2 with automatic selection
diffreport.annotated <- annotateDiffreport(xsg.fill, fc_th=4,
polarity="positive")
```

Both examples generate a data-frame, identical to the normal `diffreport` result, but now with three additional result columns from CAMERA. In example 1 we perform a quick-run, that means we only generate the `xsAnnotate` object und call `groupFWHM` and `findIsotopes`. From these results we preselect 3 pseudospectra (10,11,30), taken from the column `pc`. In the next run, we run `annotateDiffreport` again with our list as parameter. An annotation will only be done for these three groups. In example 2 we perform an automatic preselection, where the `fc_th` parameter defines a threshold for selecting groups, which contains ions with a fold change higher than four. For other pseudospectra, no adduct annotation will be calculated. The fold change value is taken from the `diffreport` result. For other parameters see the manpage of `annotateDiffreport`.

### 2.2.3 The function `annotate`

`Annotate` is a wrapper function for working with multiple samples, which have the same sample class. It's similar to `annnoteDiffreport`, but doesn't use the `diffreport`. Therefore you can't use the parameter `fc_th` and `pval_th` for an automatic selection. Only a handmade preselection with the parameter `pval_list` is possible. A "quick" mode is also available, that runs only `groupFWHM` and `findIsotopes`. The normal mode runs `groupFWHM`,

findIsotopes, groupCorr and findAdducts in order as mentioned. Every parameter of these functions also works with annotate. As a small example:

```
#A full annotation run
an <- annotate(xs, perfwHM=0.7, cor_eic_th=0.75,
ppm=10, polarity="positive")
#Generate result
peaklist <- getPeaklist(an)
#Save results
write.csv(peaklist,file="results.csv")
```

## 2.3 Interpretation of the Results

id	mz	rt	isotopes	adduct	pc
65	176.04	280.09			
76	136.05	280.43	[14][M+1]1+		5
77	135.05	280.43	[14][M]1+		5
74	153.06	280.43		[M+H]+ 152.05437	5
75	175.04	280.43		[M+Na]+ 152.05437	5
73	197.02	280.76		[M+2Na-H]+ 152.05437	5
78	377.74	286.15			
79	732.5	286.49			
83	488.32	286.82		[M+Na]+ 465.33205	7
82	466.34	286.82		[M+H]+ 465.33205	7
...					

Table 2: Example of annotation result for one sample. Columns with intensity values are omitted. blue-line: annotated group 5, red-line: annotated group 7

Table 2 shows an example of annotation results. A small cutout of the result table is displayed, the columns with the intensity values are omitted and the rows are ordered by their rt values for better readability. The column *pc* shows the result of the peak correlation based annotation (independent of the annotations *iso* and *adduct*). Peaks with the same label are supposed to belong to the same spectrum. The column *adduct* shows the annotation hypotheses for the ions. The value after the brackets is the estimated molecular mass.

The column *isotopes* contains the annotated isotopes for a monoisotopic peak. The values in the first square brackets denote the isotope-group-id(column *id*), the second is the isotope annotation and the number after the brackets is the charge of the isotope.

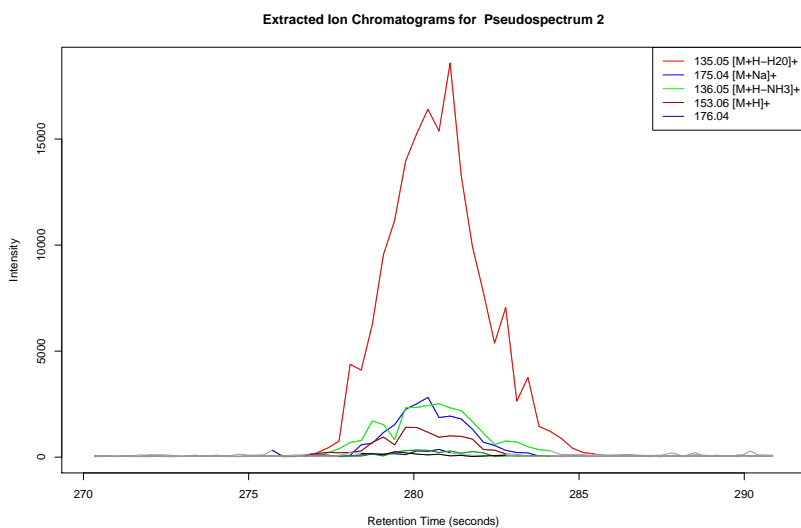


Figure 2: EICs.

## 2.4 Annotation without verification by correlation

A short notice for former *esi* user, this step is now obsolete and not longer supported. All annotations use the peak correlation if possible.

## 2.5 Visualisation of the Results

For a graphical presentation of the annotation result CAMERA provides the function `plotEICs` to visualize the raw data and the function `plotPsSpectrum` to plot all peaks of a pseudospectrum. The next examples shows the use of both functions.

```
> library(CAMERA)
> file <- system.file("mzdata/MM14.mzdata", package = "CAMERA")
> xs <- xcmsSet(file, method = "centWave", ppm = 30, peakwidth = c(5,
+   10))
> an <- xsAnnotate(xs)
> an <- groupFWHM(an)
> an <- findAdducts(an, polarity = "positive")
> plotEICs(an, pspecIdx = 2, maxlabel = 5)
```

In figure 2 you see the EICs of all peaks from one pseudospectrum. With this plot you can manual check if the grouping makes sense. In the other figure 3 you see a typical  $m/z$  plot, with labelled, annotated peaks.



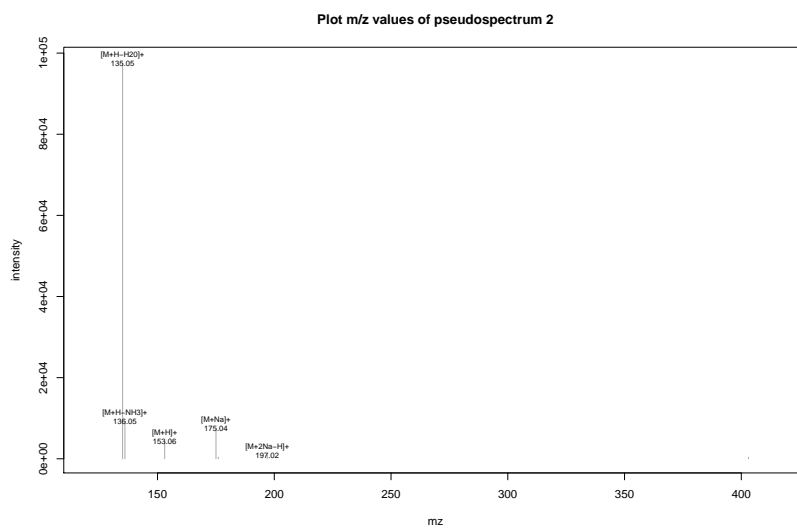


Figure 3: Spectra.

```
> plotPsSpectrum(an, pspec = 2, maxlabel = 5)
```

## References

- [1] Ralf Tautenhahn, Christoph Böttcher, Steffen Neumann : Annotation of LC/ESI-MS Mass Signals, BIRD 2007 Proc. of BIRD 2007 – 1st International Conference on Bioinformatics Research and Development, 2007. <http://www.springerlink.com/content/4731404001787974/> and <http://msbi.ipb-halle.de/~rtautenh/bird07.pdf>